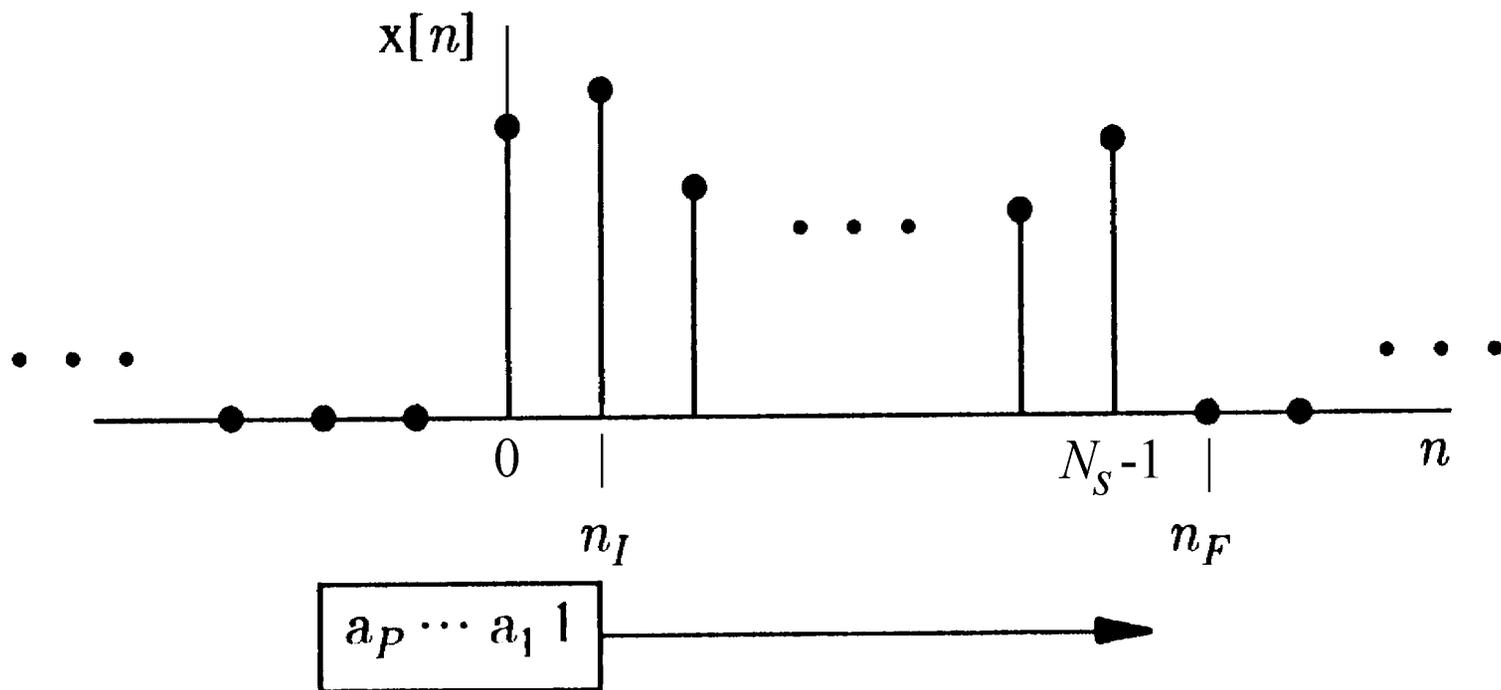


AR MODELING VIA LINEAR PREDICTION

- Linear prediction as a least squares problem
- Autocorrelation method
- Covariance method
- Modified covariance method
- Burg's method

LINEAR PREDICTION OF DATA



LEAST SQUARES LINEAR PREDICTION

LINEAR PREDICTION EQUATION FOR DATA

$$\underbrace{\begin{bmatrix} \hat{x}[n_I] \\ \hat{x}[n_I + 1] \\ \vdots \\ \hat{x}[n_F] \end{bmatrix}}_{\hat{\mathbf{x}}_0} = - \underbrace{\begin{bmatrix} x[n_I - 1] & x[n_I - 2] & \cdots & x[n_I - P] \\ x[n_I] & x[n_I - 1] & \cdots & x[n_I - P + 1] \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ x[n_F - 1] & x[n_F - 2] & \cdots & x[n_F - P] \end{bmatrix}}_{\mathbf{X}_1} \underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_P \end{bmatrix}}_{\mathbf{a}'}$$

LEAST SQUARES FORMULATION

$$-\mathbf{X}_1 \mathbf{a}' \stackrel{\text{ls}}{=} \mathbf{x}_0; \quad \mathbf{a}' = -\mathbf{X}_1^+ \mathbf{x}_0$$

YULE-WALKER EQUATIONS

The linear prediction equations can be written as

$$-\mathbf{X}_1 \mathbf{a}' = \mathbf{x}_0 - \epsilon \quad \text{or} \quad \begin{bmatrix} | & & | \\ \mathbf{x}_0 & \mathbf{X}_1 & \\ | & & | \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{a}' \end{bmatrix} = \mathbf{X} \mathbf{a} = \epsilon$$

The Orthogonality Theorem can be expressed as

$$\mathbf{X}^{*T} \epsilon = \begin{bmatrix} \text{---} & \mathbf{x}_0^{*T} & \text{---} \\ & \mathbf{X}_1^{*T} & \end{bmatrix} \epsilon = \begin{bmatrix} \mathcal{S} \\ \mathbf{0} \end{bmatrix}$$

The combination of these produces

$(\mathbf{X}^{*T} \mathbf{X}) \mathbf{a} = \begin{bmatrix} \mathcal{S} \\ \mathbf{0} \end{bmatrix}$	least squares Yule-Walker
---	------------------------------

METHODS OF LINEAR PREDICTION (cont'd.)

EQUATION FOR PREDICTION ERROR

$$\underbrace{\begin{bmatrix} \epsilon[n_I] \\ \epsilon[n_I + 1] \\ \vdots \\ \vdots \\ \epsilon[n_F] \end{bmatrix}}_{\boldsymbol{\epsilon}} = \underbrace{\begin{bmatrix} X[n_I] & X[n_I - 1] & \cdots & X[n_I - P] \\ X[n_I + 1] & X[n_I] & \cdots & X[n_I - P + 1] \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ X[n_F] & X[n_F - 1] & \cdots & X[n_F - P] \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_P \end{bmatrix}}_{\mathbf{a}}$$

LEAST SQUARES YULE-WALKER EQUATIONS

$$(\mathbf{X}^{*T} \mathbf{X}) \mathbf{a} = \begin{bmatrix} \mathcal{S} \\ \mathbf{0} \end{bmatrix} \quad \text{where } \mathcal{S} = \min \|\boldsymbol{\epsilon}\|^2$$

COMPARISON OF DATA MATRICES

**AUTOCORRELATION
METHOD: MATRIX X**

$$\begin{bmatrix} x[0] & 0 & \cdots & 0 \\ x[1] & x[0] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x[P] & x[P-1] & \cdots & x[0] \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ x[N_s-1] & x[N_s-2] & \cdots & x[N_s-P-1] \\ 0 & x[N_s-1] & \cdots & x[N_s-P] \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x[N_s-1] \end{bmatrix}$$

**COVARIANCE
METHOD: MATRIX X**

$$\begin{bmatrix} x[P] & x[P-1] & \cdots & x[0] \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ x[N_s-1] & x[N_s-2] & \cdots & x[N_s-P-1] \end{bmatrix}$$

COMPARISON OF METHODS

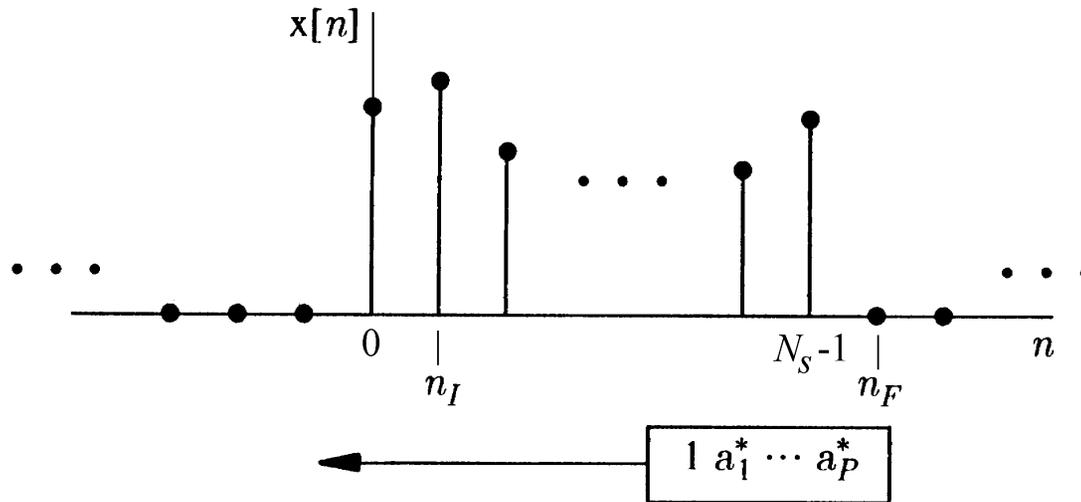
AUTOCORRELATION METHOD

- Correlation matrix $\mathbf{X}^{*T}\mathbf{X}$ is Toeplitz
- Filter guaranteed to be minimum-phase

COVARIANCE METHOD

- Generally a lower average prediction error
- AR model *could* be unstable

BACKWARD LINEAR PREDICTION



- Leads to equations $\tilde{\epsilon}^b = \tilde{\mathbf{X}}\mathbf{a}^*$ and $(\tilde{\mathbf{X}}^{*T}\tilde{\mathbf{X}})\mathbf{a}^* = \begin{bmatrix} \mathcal{S}^b \\ \mathbf{0} \end{bmatrix}$.

MODIFIED COVARIANCE METHOD

- Minimizes the sum of forward *and* backward squared errors.
- Leads to least squares Yule-Walker equations of the form

$$\left(\mathbf{X}^{*T}\mathbf{X} + \tilde{\mathbf{X}}^T\tilde{\mathbf{X}}^*\right) \mathbf{a} = \begin{bmatrix} \mathcal{S}^{fb} \\ \mathbf{0} \end{bmatrix}$$

where \mathcal{S}^{fb} is the sum of forward and backward squared errors.

MODIFIED COVARIANCE METHOD (cont'd.)

This method seeks to minimize

$$\mathcal{S}^{fb} = \sum_{n=P}^{N_s-1} (|\epsilon[n]|^2 + |\epsilon^b[n]|^2) = \|\epsilon\|^2 + \|\tilde{\epsilon}^b\|^2$$

where

$$\|\epsilon\|^2 = \epsilon^{*T} \epsilon = \mathbf{a}^{*T} \mathbf{X}^{*T} \mathbf{X} \mathbf{a} \quad \text{and} \quad \|\tilde{\epsilon}^b\|^2 = \mathbf{a}^T \tilde{\mathbf{X}}^{*T} \tilde{\mathbf{X}} \mathbf{a}^* = \mathbf{a}^{*T} \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}^* \mathbf{a}$$

The problem is therefore to minimize

$$\mathcal{S}^{fb} = \mathbf{a}^{*T} \left(\mathbf{X}^{*T} \mathbf{X} + \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}^* \right) \mathbf{a}$$

subject to the constraint

$$\mathbf{a}^{*T} \boldsymbol{\iota} = 1 \quad \text{where} \quad \boldsymbol{\iota} = [1 \ 0 \ \dots \ 0]^T$$

MODIFIED COVARIANCE METHOD (cont'd.)

Use Lagrange multiplier and complex gradient to obtain:

$$\begin{aligned}\nabla_{\mathbf{a}^*} \left[\mathbf{a}^{*T} \left(\mathbf{X}^{*T} \mathbf{X} + \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}^* \right) \mathbf{a} + \lambda (1 - \mathbf{a}^{*T} \boldsymbol{\iota}) + \lambda^* (1 - \boldsymbol{\iota}^T \mathbf{a}) \right] \\ = \left(\mathbf{X}^{*T} \mathbf{X} + \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}^* \right) \mathbf{a} - \lambda \boldsymbol{\iota} = \mathbf{0}\end{aligned}$$

Multiply by \mathbf{a}^{*T} and observe:

$$\underbrace{\mathbf{a}^{*T} \left(\mathbf{X}^{*T} \mathbf{X} + \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}^* \right) \mathbf{a}}_{\mathcal{S}^{fb}} - \lambda \underbrace{\mathbf{a}^{*T} \boldsymbol{\iota}}_1 = 0 \quad \implies \quad \lambda = \mathcal{S}^{fb}$$

Substitute $\lambda = \mathcal{S}^{fb}$ in the top equation to find:

$$\boxed{\left(\mathbf{X}^{*T} \mathbf{X} + \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}^* \right) \mathbf{a} = \begin{bmatrix} \mathcal{S}^{fb} \\ \mathbf{0} \end{bmatrix}}$$

BURG'S METHOD

- Estimates parameters in the *lattice* form of the filter:

$$\hat{\gamma}_p = \frac{2 \sum_n \epsilon_{p-1}[n] \epsilon_{p-1}^b[n-1]}{\sum_n (|\epsilon_{p-1}[n]|^2 + |\epsilon_{p-1}^b[n-1]|^2)}$$

- Minimizes the sum of forward and backward squared errors.

DERIVATION OF BURG'S METHOD

This method applies the lattice form of the PEF to the data to write:

$$\epsilon_p = \begin{bmatrix} \epsilon_p[p] \\ \epsilon_p[p+1] \\ \vdots \\ \epsilon_p[N_s - 1] \end{bmatrix} = \begin{bmatrix} \epsilon_{p-1}[p] \\ \epsilon_{p-1}[p+1] \\ \vdots \\ \epsilon_{p-1}[N_s - 1] \end{bmatrix} - \gamma_p^* \begin{bmatrix} \epsilon_{p-1}^b[p-1] \\ \epsilon_{p-1}^b[p] \\ \vdots \\ \epsilon_{p-1}^b[N_s - 2] \end{bmatrix}$$

and

$$\epsilon_p^b = \begin{bmatrix} \epsilon_p^b[p] \\ \epsilon_p^b[p+1] \\ \vdots \\ \epsilon_p^b[N_s - 1] \end{bmatrix} = \begin{bmatrix} \epsilon_{p-1}^b[p-1] \\ \epsilon_{p-1}^b[p] \\ \vdots \\ \epsilon_{p-1}^b[N_s - 2] \end{bmatrix} - \gamma_p'^* \begin{bmatrix} \epsilon_{p-1}[p] \\ \epsilon_{p-1}[p+1] \\ \vdots \\ \epsilon_{p-1}[N_s - 1] \end{bmatrix}$$

BURG'S METHOD (cont'd.)

Note the structure of these equations:

$$\begin{bmatrix} \times \\ \text{---} \\ \mathbf{e}_p^f \end{bmatrix} = \begin{bmatrix} \epsilon_p[p] \\ \text{---} \\ \epsilon_p[p+1] \\ \vdots \\ \epsilon_p[N_s-1] \end{bmatrix} = \underbrace{\begin{bmatrix} \epsilon_{p-1}[p] \\ \epsilon_{p-1}[p+1] \\ \vdots \\ \epsilon_{p-1}[N_s-1] \end{bmatrix}}_{\mathbf{e}_{p-1}^f} - \gamma_p^* \underbrace{\begin{bmatrix} \epsilon_{p-1}^b[p-1] \\ \epsilon_{p-1}^b[p] \\ \vdots \\ \epsilon_{p-1}^b[N_s-2] \end{bmatrix}}_{\mathbf{e}_{p-1}^b}$$

$$\begin{bmatrix} \mathbf{e}_p^b \\ \text{---} \\ \times \end{bmatrix} = \begin{bmatrix} \epsilon_p^b[p] \\ \epsilon_p^b[p+1] \\ \vdots \\ \text{---} \\ \epsilon_p^b[N_s-1] \end{bmatrix} = \underbrace{\begin{bmatrix} \epsilon_{p-1}^b[p-1] \\ \epsilon_{p-1}^b[p] \\ \vdots \\ \epsilon_{p-1}^b[N_s-2] \end{bmatrix}}_{\mathbf{e}_{p-1}^b} - \gamma_p'^* \underbrace{\begin{bmatrix} \epsilon_{p-1}[p] \\ \epsilon_{p-1}[p+1] \\ \vdots \\ \epsilon_{p-1}[N_s-1] \end{bmatrix}}_{\mathbf{e}_{p-1}^f}$$

BURG'S METHOD (cont'd.)

Choose γ_p to minimize:

$$\begin{aligned}\mathcal{S}_p^{fb} &= \|\epsilon_p\|^2 + \|\epsilon_p^b\|^2 \\ &= (\mathbf{e}_{p-1}^f - \gamma_p^* \mathbf{e}_{p-1}^b)^{*T} (\mathbf{e}_{p-1}^f - \gamma_p^* \mathbf{e}_{p-1}^b) + (\mathbf{e}_{p-1}^b - \gamma_p \mathbf{e}_{p-1}^f)^{*T} (\mathbf{e}_{p-1}^b - \gamma_p \mathbf{e}_{p-1}^f) \\ &= (1 + \gamma_p \gamma_p^*) (\|\mathbf{e}_{p-1}^f\|^2 + \|\mathbf{e}_{p-1}^b\|^2) - 2\gamma_p^* (\mathbf{e}_{p-1}^f)^{*T} \mathbf{e}_{p-1}^b - 2\gamma_p (\mathbf{e}_{p-1}^b)^{*T} \mathbf{e}_{p-1}^f\end{aligned}$$

The minimization condition (using the complex gradient) is

$$\nabla_{\gamma_p^*} \mathcal{S}_p^{fb} = \gamma_p (\|\mathbf{e}_{p-1}^f\|^2 + \|\mathbf{e}_{p-1}^b\|^2) - 2(\mathbf{e}_{p-1}^f)^{*T} \mathbf{e}_{p-1}^b = 0$$

or

$$\gamma_p = \frac{2(\mathbf{e}_{p-1}^f)^{*T} \mathbf{e}_{p-1}^b}{\|\mathbf{e}_{p-1}^f\|^2 + \|\mathbf{e}_{p-1}^b\|^2}$$

BURG'S METHOD SUMMARY

INITIALIZATION

$$\begin{bmatrix} \times \\ \text{---} \\ e_0^f \end{bmatrix} = \begin{bmatrix} e_0^b \\ \text{---} \\ \times \end{bmatrix} = \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N_s - 1] \end{bmatrix}$$

RECURSION

$$\gamma_p = \frac{2(e_{p-1}^f)^* e_{p-1}^b}{\|e_{p-1}^f\|^2 + \|e_{p-1}^b\|^2}$$

$$\begin{bmatrix} \times \\ \text{---} \\ e_p^f \end{bmatrix} = e_{p-1}^f - \gamma_p^* e_{p-1}^b ; \quad \begin{bmatrix} e_p^b \\ \text{---} \\ \times \end{bmatrix} = e_{p-1}^b - \gamma_p e_{p-1}^f$$

BURG'S METHOD IN MATLAB

MAIN LOOP

```
L=length(ef);  
for p=1:P;  
    gamma(p) = (2*ef'*eb)/(ef'*ef + eb'*eb)  
    if p < P  
        tmp1 = ef - conj(gamma(p))*eb;  
        tmp2 = eb - gamma(p)*ef;  
        ef = tmp1(2:L);  
        eb = tmp2(1:L-1);  
    end;  
    L=L-1;  
end;
```

AR ORDER SELECTION

THEORETICAL CRITERIA

AIC $P = \operatorname{argmin} N_s \ln \sigma_{\varepsilon_P}^2 + 2P$

CAT $P = \operatorname{argmin} \left(\frac{1}{N_s} \sum_{p=1}^P \frac{N_s - p}{N_s \sigma_{\varepsilon_p}^2} \right) - \frac{N_s - P}{N_s \sigma_{\varepsilon_P}^2}$

FPE $P = \operatorname{argmin} \sigma_{\varepsilon_P}^2 \left(\frac{N_s + P + 1}{N_s - P - 1} \right)$

MDL $P = \operatorname{argmin} N_s \ln \sigma_{\varepsilon_P}^2 + P \ln N_s$

AR ORDER SELECTION (cont'd.)

EFFECTIVE RANK METHOD FOR DATA MATRICES

1. Choose \mathbf{X} with column dimension larger than the suspected order and perform the SVD.
2. If r is the rank of \mathbf{X} find r' such that the singular values satisfy (approximately)

$$\left(\frac{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_{r'}^2}{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2} \right)^{\frac{1}{2}} \approx 1$$

3. Choose P to be one less than the effective rank, $P = r' - 1$.